A version of the paper has been published in:
International Journal of Production Research, volume 50, issue 13, pp. 3579-3593, 2012.

Available in: http://www.tandfonline.com/toc/tprs20/50/13

Evaluation by simulation to optimize information systems personalization quality in logistics

Makram SOUI*[*], Mourad Abed*, Christophe Kolski* and Khaled Ghèdira**

**Univ Lille of France, F-59000 Lille, France,UVHC, LAMIH, F-59313 Valenciennes, France CNRS, UMR 8530, F-59313 Valenciennes, France.*

***Institute of Management of Tunis, 41, Rue de la liberté, Cité Bouchoucha, Bardo 2000, Tunisia.*

Personalized Information Systems (PIS), related to different fields (travelling and mobility, production, logistics…), represent an object of many research and development perspectives. Using PIS, it becomes possible to supply the user only with the pertinent information that suits his/her preferences. In deed, thanks to personalization, user may feel that PIS is developed for him/her. This system adaptation becomes a necessity for the user's satisfaction. In this context, many studies were oriented toward the user modeling, the design methods of PIS and the personalization algorithms, etc., but, the evaluation of these systems is neglected. Difficulties concerning context-centred evaluation appear. This article is focussed on the evaluation of the personalized information systems in order to optimize the personalization quality according to several criteria. For such systems, it is important to envisage new adapted evaluation methods. An evaluation method using simulation of a model of PIS, called MetSim (Method evaluation per Simulation) is proposed. MetSim is also based on Case Based Reasoning system to identify problems. This evaluation approach has been validated by applying it to assess PIS in logistics field.

Keywords: Personalized information system, Evaluation, Simulation, Personalization, User interface (UI).

1. Introduction

Nowadays, we are living with a very fast evolution of the personalized information systems (PIS) (Karat *et al*. 2004) (Fan and Poole 2006) (Piller and Tseng 2010); such systems can be used in everyday life and/or at work, in many different contexts; production (Sugimori *et al.*

[*] Corresponding author. Email: Souii_makram@yahoo.fr

2007) (Payne and Cariapa 2000), logistics (Chu-hua *et al.* 2011), transport are domains in which personalization principles offer new perspectives (Chabrol *et al*. 2006).

In general, these systems use basic information about the user, the categories of users, and/or an analysis of contextual data (historical information) to identify the preferences, profile, and characteristics of the user. PIS should provide the user with relevant information taking into account the context when using the system. In fact, personalization allows adapting the human-machine interaction with regard to the context of use which must be considered when designing and executing the system. Three categories of contextual information are often distinguished in the literature: information about the user, the platform used and the environment. Indeed, many works are carried out both in academic world and in companies about user modelling, methods for designing PIS, tools for the generation and management of personalized web sites and personalization algorithms; see for instance (Houben et al., 2009). However, the evaluation of personalized systems is often neglected; very few methods for PIS evaluation exist.

With a view of PIS evaluation, it is important to consider explicitly the context as central element of evaluation. Thus, various contexts of use (P: platform, U: user, and E: environment) have to be taken into consideration.

We propose an evaluation method using simulation. Dedicated to PIS, this method is entitled MetSim (Method evaluation by SIMulation). First, we present a state of art about the notions of personalization and evaluation. Then, we list some existing evaluation methods. We describe also a simulation approach for PIS evaluation. This approach is illustrated in the logistics domain. A conclusion and prospects end this article.

2. What is personalization?

The personalized information system represents an important potential and is object of many research and development perspectives (Sancho 2005) (MacComascaigh and Andrews 2009) (De Bra 2010). The aim of PIS is to provide the user with the needed information taking into account his/her preferences (Basu *et al*. 1998) (Sancho 2005). Logistics is a rich and complex domain because of diversity of activities and contexts of use. In this context, personalization may offer new perspectives. For example, responsible for supply chain can hope to have at their disposal only some information according to his/her activity (production, distribution,…) just what they are directly interested in. Indeed, they are meant to have access to a reliable, multi-modal and personalized information using various supports (PC, PDA, smart phone, tablet, etc.) (Zimmermann and Lorenz 2008). The technical capabilities of these devices allow for information to be retrieved from or delivered to nearly everywhere (at manufacturing plant, at assembly plant, at distribution centres,…).

Numerous systems allow the personalization of the interaction between user and system. As illustration we may first mention Letizia (Lieberman 1995): it is an intelligent system which helps users in web navigation. It implicitly sets the user profile from the pages content he/she explored. Then it looks for other interesting pages and presents them in an independent window. Pages selection is based on a similarity score between the pages content and the user profile. Web personae (Gowan 2003) is a system that interacts in an off line mode with web application. It is composed of a constructor and an identifier. The constructor progressively sets the user model which is a list of profiles corresponding to user preferences given from the web interactions. The identifier finds out the user profile which is related to the system current use. It is based on collaborative filtering to provide personalized information. InfoSleuth (Bayardo 1997) (Nodine *et al*. 2003) is a system dedicated for search of information in dynamic, heterogeneous and distributed environment. The InfoSleuth architecture includes different agents' models (User, Ontology, Broker, Ressource, Data

Analysis, Task Execution, Query and Monitor) which communicate and cooperate for the personalization. Syskill & webert (Pazzani *et al*. 1996) is a recommendation system that helps the user by offering him/her relevant pages links. This system presents the user profile under the form of clusters. Each one corresponds to a user preference which is modelled according to a boolean vector of key words. The selection of recommended pages is based on a probability calculation.

In general, we can say that personalization deals with the capacity of adaptation of a system considering some information related to the context of use (P: platform, U: user, and E: environment).

3. The evaluation: definitions and principles

Currently, there is no standard definition, in the field of research, of the concept of the interactive system evaluation. Thus, according to (Huart and al. 2008) evaluation of an interactive system includes both its verification and its validity. The system is verified once it corresponds to the specific needs defined by the designer. It is validated when it corresponds to the needs regarding the application domain constraints. As for (Senach 1990), "evaluation consists in comparing a model of evaluated object to a referential model permitting to draw conclusions". According to (Le Bodic 2005), the evaluation consists in the validation of adequacy between functional realization and the scenario of usage. Real experiments playing these scenarios will be accomplished to assess the usability of information system.

In logistics context, there are many situations in which evaluation may be helpful. In fact, a large number of models have been presented that aim at finding the optimal system. We may mention (Jingshan and Ningjian 2007): a simple Markovian model to evaluate the quality performance of a flexible manufacturing system. The analytical hierarchy process (AHP) and the analytical network process (ANP) (Yang et al. 2011): are utilized to best coordinate buyer and supplier efforts to improve productivity and competitive advantage in a supply chain. The SCOR model (Supply Chain Operations References) (Huang et al. 2005) (Francis 2007), which consists of comparing the performance of a logistics system to a 'best-practice' standard. The SCOR model makes use of several performance parameters that range from highly aggregated indicators (named key performance indicators (KPIs)) to indicators describing a specific operational issue.

Concerning us, we are interested in PIS evaluation; evaluation consists in ensuring that the system provides the user with relevant information. This evaluation may be done on two levels:

- At the level of the contents and the services; evaluation consists in ensuring that predictions proposed by the system correspond to user preferences.
- At the level of the user interface; it may cope with user preferences related to the container, in addition to the context in which the application is used. The evaluation of the interface consists in checking its capacity of dynamic adaptation (change of the style of display, change of the color of background, etc.) during the contextual change.

4. Methods for interactive system evaluation

There are various methods for the evaluation of interactive system (Askin et al. 1999) (Ivory, Hearst 2001) (Sears and Jacko, 2008) (Lapena *et al.* 2008) (Janssen *et al.* 2010).To present the existing evaluation methods, we will keep the classification of (Huart et al. 2008) inspired by the one proposed in (Whitefield 1991).

4.1 Empirical approaches

The empirical evaluation consists in collecting behavioural information concerning the system use. This type of evaluation necessitates the presence of user and the existence of a version of the system (model, mock-up, prototype or final system). The use of a system is observed and analysed. Among the many empirical methods which exist, we can cite: interviews, questionnaires (Corlett et al. 1995), analysis of traces of use (Tarby et al. 2009), electronic informer (Drury 1990), eye-tracking (Duchowski 2003).

These methods have been used since many years to assess the usability of information system according to ergonomic criteria (ISO 2006). However, the personnalization criteria are neglected. These approaches require adaptations to be able to support the evaluation of PIS. For instance, the use of interviews by (Ganneau et al. 2008) to evaluate EMMA system (Embedded Manager for Mobile Adaptation) presents a significant example of the prospects offered by this technique. In addition we have proposed in (Soui et al. 2007) a questionnaire to discover the problems related to the personalization criteria during the evaluation of MouverPerso system.

4.2 Expert qualified approaches

This type of evaluation based on the judgement of an expert in ergonomics or a specialist in human-machine interaction. It compares the performances and characteristics which are presented in the form of specifications, model or of prototype, to the standards or recommendations in order to detect a design errors. Among the many expert qualified methods that exist, we can cite: specialist intervention, heuristic evaluation (Nielsen and Molich 1990), cognitive walkthrough (Mahatody et al., 2010), guideline reviews (Vanderdonckt 1999) (Bereikdar 2004).

These methods contribute to improve the ergonomic characteristics of the user interface. But, a few propositions and studies in the literature were set about the PIS evaluation. For example, (Magoulas et al. 2003) proposed integrating heuristic evaluation for adaptive learning environments. In fact, they modified the Nielsen's heuristics (Nielsen 1993) to take in consideration the specificities of personalization.

4.3 Analytical methods

This category of method allows assessing the system design and not its use. Analytical modeling is engineering approach to usability evaluation that enables evaluators to predict usability with user and interface models. A wide range of modeling techniques has been developed and they support different types of analyses. We can classify analytical approaches into the following three categories: (1) the predictive models that are based on breaking down the tasks of potential users into users'actions and users'cognitive processes, (2) the automatic evaluation tools which analyze and verify the UI code and (3) the simulation methods (Feng et al. 2009). This simulation allows evaluator to assess hypotheses in term of models and scenario and so to guide, either system evaluation or design. So far, there has been little experience in the application of analytical approaches to adaptive systems. These methods are oriented towards usability problems concerning interactive systems. In fact, the predictive models do not take into account the specificity of PIS. Moreover, the capture techniques of user behaviours by automatic tools seem to be promising in the evaluation of PIS. (Le bodic et al. 2005) proposed also a simulator based on a computing simulation of a virtual user to evaluate multimodal user interfaces. This simulator was built on three models: environment, user and artefact description.

We notice that there are many methods contributing to the evaluation of interactive systems. To our knowledge, there is not specific method devoted to PIS evaluation. The current methods are in fact firstly oriented towards the usability evaluation; but we think that some of them can be adapted for PIS evaluation.

In our work, we distinguish two evaluation levels:

-The first level consists in evaluating the learning capacity of the PIS. It concerns mainly the evaluation of personalization related to the content and is carried out during the system execution. This evaluation allows verifying if the personalization offered by the system is able to take into account the user preferences. For example, for the production system, when a responsible of distribution chooses an itinerary (planning) among others, his/her choice is based on a certain criteria of selection. The distribution way is the path followed by a product or service, from the stage of production to consumption one. When evaluating itineraries, we have to sort them out according to different criteria. Thus, the PIS assign ranks for each itinerary. We may define (itinerary) by a set of values that corresponds to its ranks via different criteria (k is a number of criteria the user is concerned with): $S_i = \{v_1, v_2, v_3,.., v_k\}$. With v_i corresponds to the ranks of solutions according to the criteria C_i, $v_i = rank(C_i)$, $1<v_i<n$. with n is the number of itineraries. Consider two values v_i and v_j of the criterion C_k related to the two itineraries respectively S_i and S_j : $v_{ki} < v_{kj} \Leftrightarrow S_i \textit{ is preferred to } S_j, \textit{ for the criterion } C_k$. For the evaluation of personalization related to the content, we suggest a module of evaluation based on multi-agent simulation (user agent, criteria agent, evaluator agent). The idea behind the proposed model is to solve the problem of contradictory criteria. For this reason, we associate a criterion agent for each criterion. Thus, criterion agents cooperate in order to generate the optimal solution. To come to a compromise between the different criterion agents, we propose a negotiation process. In fact, to respond to the user demand, the user agent transfers his/her request to personalized system that will look for the possible itineraries. These solutions will be examined then by criterion agents in order to select the preferable one. Thus, the evaluator agent verifies the system quality in regard to the proposed itineraries (by system) and the actually chosen (expected) one by the user.

The second level consists in verifying the PIS reaction after the context changes. The PIS must have the capacity to be adapted to its context of use by discovering and using contextual information such as the localization of the user, the characteristics of the target platform, the environmental conditions, etc. Thus, various contexts of use (P: platform, U: user, and E: environment) are taken into consideration for PIS evaluation. Therefore we have to focus on these contexts to assess these systems (Zimmermann and Lorenz 2008) (Norsham and al. 2009). This level concerns the evalution related to the container of PIS. This article details the second level of evalution.

5. PIS Simulation: Evaluation Method

The suggested method consists in evaluating a PIS while varying hypothesis of use related to different contexts. The aim of simulation is to be able to predict various scenarios by changing the context of use (Platform, User, and environment). In our method the Human-Machine Interaction (HMI) is depicted as a set of components. This HMI is represented in UsiXML (USer eXtensible Interface Markup Language) (Limbourg *et al*. 2004). Thanks to the language used to describe the HMI abstraction, it is possible to have an access to this description and to alter it during the design.

This evaluation method will be presented through IDEF0 technique (Ang 1999). This method is composed of three phases (Soui *et al*. 2010): (1) Proposal of the hypothesis (Figure 1, A1), (2) Identification of the problems (Figure 1, A2), (3) Analysis of the results (Figure 1, A3). Each phase will be presented with its detailed activities.

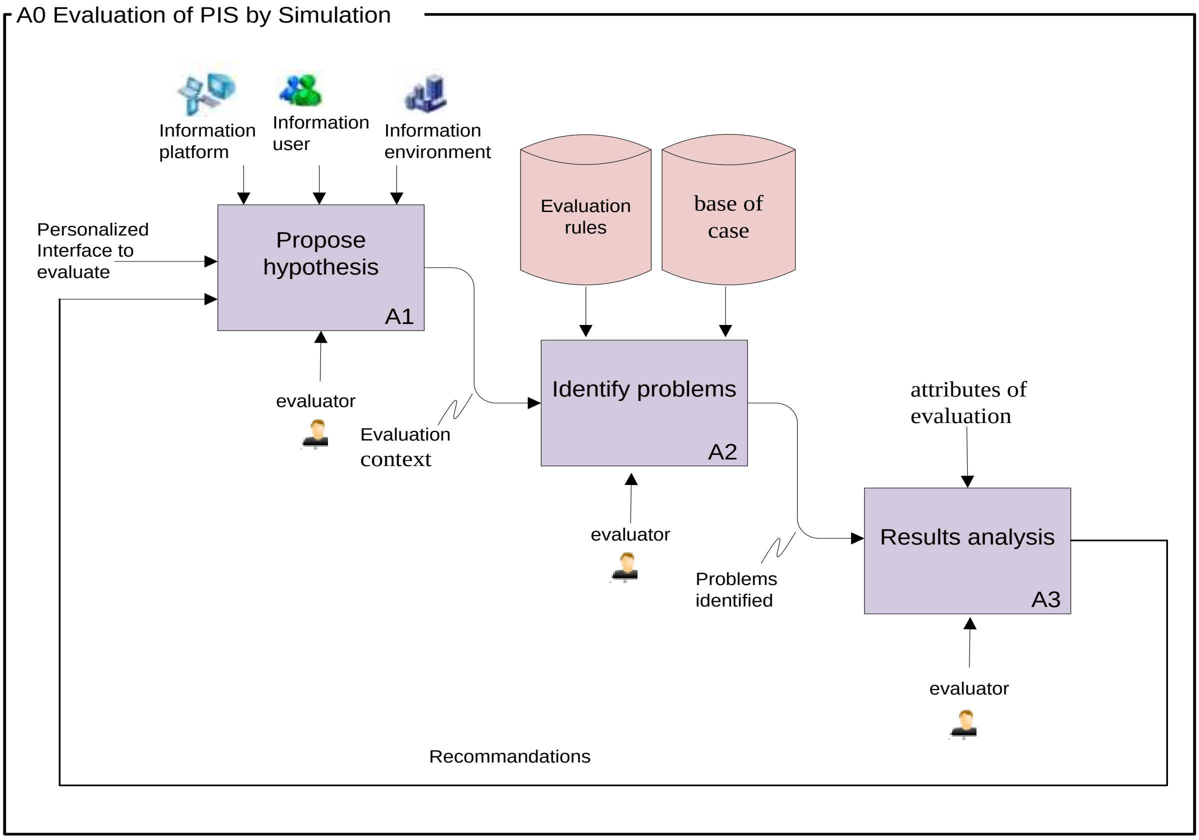

Figure 1. Phases of MetSim method.

5.1 Proposal of the hypothesis

Figure 2 presents the activity A1 "propose hypothesis". The proposal starts with the definition of the context of evaluation. This definition has as an in input the user interface to be evaluated. The latter is described in UsiXML.

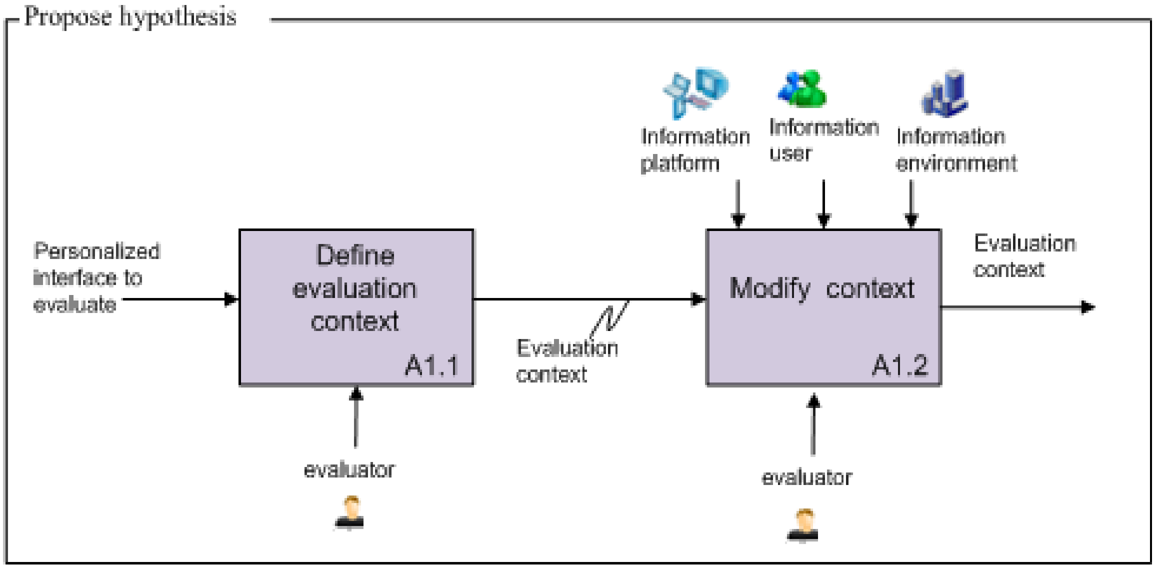

Figure 2. Proposal of the hypothesis.

A1.1 activity consists in defining the context of evaluation. Thus, according to the criterion that must be checked, the evaluator must choose the category of the subjects which will be implied in evaluation (beginners, experts, etc.), the support with which the subjects interact with the system during the evaluation (PDA, PC, etc.), and the environment of the interaction.

Then, the evaluator can modify the tag <contextModel> which includes the contexts of use envisaged according to the triplet <platform, user, environment>. Thanks to the language used to describe the HMI, it is possible to reach this part and to modify it during the system design.

5.2 Identification of the problems

The phase A2 permits to detect the problems caused due to the context change basing on a Case Based Reasoning (C.B.R.) system (Kolodner 1993). The principle of MetSim consists in treating a new case (case targets) by counting on previous former experiments (case of reference). This type of reasoning rests on the following hypothesis: if a past experiment and the new situation are sufficiently similar, then all that can be explained or applied to the past experiments (case bases) remains valid if one applies it to the new situation which represents the new problem to be solved (Bradley *et al*. 2000) (Coyle and Cunninghan 2002) (Lajmi *et al*. 2009). The CBR is subdivided in several connected stages forming a cycle presented on figure 3.

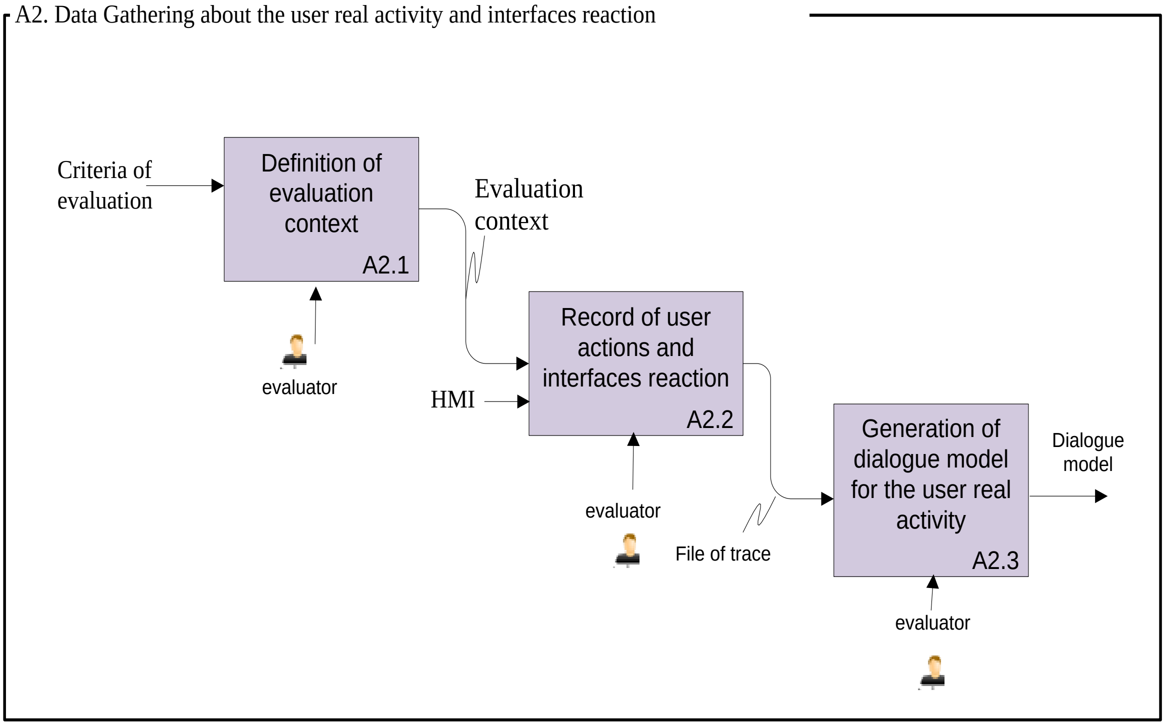

Figure 3. Identification of the encountered problems.

The cases are real experiments that are introduced in the form: triplet <R, P, J> where:
- R is the description of the reaction of the user interface
- P is the description of the problem
- J is the analysis of the problem

To be usable, a base of case must contain a certain number of cases. An empty base of case does not allow any reasoning. Consequently, it is important to initiate the base of case with relevant cases. In our work, the base of case is initiated based on a set of rules of evaluation. These rules are initially defined by the evaluator. Indeed, with each criterion of assessment (for example related to the environment), a number of attributes of evaluation is associated (noise, luminosity, temperature). With each attribute a unit with rules of evaluation is associated. The aim of these rules is to propose and describe the reactions of interface awaited by the user following a contextual change. Thus, the evaluator checks whether these rules were respected or not. In our method, the rule of evaluation takes the form "Action→Réaction": this action indicates a change which appears within the framework of the use (a state of the context of use to be considered by the HMI), a reaction is consequently carried out of this state.

Figure 4 presents an example of a rule of evaluation described in XML. This rule presents the reaction envisaged of the interface (bottom of screen put in gray) when the luminosity reaches the rate of 80%.

```xml
<rule> // evaluation rule (Luminosity reaches more than 80%)
  <id>R1</id>
  <action>
    <luminosity> // Luminosity reached more than 80%
    <morethan>80</morethan>
   </luminosity>
  </action>
  <reaction> // background of HMI : in gray
    <background>
      <color> clGray </color>
    </background>
  </reaction>
</rule>
```

Figure 4. Example of an evaluation rule described in XML.

A2.1 activity presents the stage of identification of the components following the adaptation process. Let us recall here that the component of HMI can have several presentations usable in various contexts. In case of change at the level of the context, an adaptation is launched. Thus, the adequate components of presentation to the context of use are arranged in containers.

The identification of the components of presentation and the containers that arrange them based on the tag <cuiModel>, which is composed of concrete containers <Concrete container>, which are attached to the main tasks (for example, a concrete container is defined for the task of search for preferred itinerary). The containers can include concrete components associated with the sub-tasks (for example, a concrete component is defined for the sub-task of the starting point). In this stage, the values of the characteristics of the containers and components reacting (such as the change of the background color, the change of the size of a window, etc.) and the reorganization of presentations are detected. The identification of the containers and components reaction following the contextual change and their characteristics are done by comparing the concrete specification of the user interface before and after the adaptation. Each adapted component or container is described in UsiXML (see figure 5).

```xml
    <comment>Generated by GrafiXML 1.2.0 build id : 200612141408</comment>
    <comment>WARNING : AUI Model save is a work in progress. Use it at your own risk</co
</head>
<auiModel id="projet1-aui_30" name="projet1-aui"/>
<cuiModel id="projet1-cui_30" name="projet1-cui">
    <window id="window_component_0" name="window_component_0"
        width="200" height="150">
        <gridBagBox id="grid_bag_box_1" name="grid_bag_box_1"
            gridHeight="17" gridWidth="20">
            <constraint gridx="10" gridy="2" gridwidth="6"
                gridheight="1" weightx="1.0" weighty="1.0"
                fill="both" insets="0,0,0,0">
                <inputText id="input_text_component_2"
                    name="input_text_component_2" isVisible="true"
                    isEnabled="true" textColor="#000000"
                    maxLength="50" numberOfColumns="15" isEditable="true"/>
            </constraint>
            <constraint gridx="2" gridy="2" gridwidth="4"
                gridheight="1" weightx="1.0" weighty="1.0"
                fill="both" insets="0,0,0,0">
                <outputText id="output_text_component_4"
                    name="output_text_component_4"
                    content="/uiModel/resourceModel/cioRef[@cioId='output_text_component
                    defaultContent="point de départ"
                    isVisible="true" isEnabled="true"
                    fgColor="#000000" isBold="true" textColor="#000000"/>
            </constraint>
        </gridBagBox>
```

Figure 5. Example of a container described in UsiXML.

Then, the system uses these characteristics extracted in order to compare the reactions of HMI of the new case with those of the former cases. For that, a search (retrieve) is carried out in case of base in order to find the similar cases.
In order to compare the new case with former ones, we need to extract the characteristics of the contents of the cases. Kolodner (1993) defines a characteristic or a descriptor by a unit (attribute, value) associated with a case. In this same view, for our approach, each case is described by a set of characteristics of the interface (For example, size of letters) and a set of attributes of context usage.

A2.2 activity presents the stage of search of the relevant case. At this level the most pertinent case is determined when using the calculating methods of similarity in order to guide research. The general algorithm of research occurs in several stages.

1. A prefiltering of the base of case provides a set of cases B approximately similar to the new one. This process of prefiltering is based on the attributes of context.
2. In parallel, for each case $C \in B$,
(a) Determine the common attributes of the reaction of the new case RN and that of the former ones RA. Let A be the set of these attributes.
(b) To calculate, for each attribute $a \in A$, the similarity Da (RNa, RAa), with RNa represents the value of the attribute a in RN and RAa, the value in RA.

$$D_a(RN_a, RA_a) = \frac{RN_a - RA_a}{DM} \in [0..1] \qquad (1)$$

DM is the difference between the high terminal and the low terminal of the domain of definition of the attribute
(c) To determine the total degree of similarity S (RN, RA)

$$S(RN, RA) = \sum_{a=1}^{A} D_a(RN_a, RA_a) \qquad (2)$$

3. Return the most similar case.

The problem is extracted from the most relevant case so as to reuse it. The latter becomes a suggested problem. It submits a process of revision and adaptation in order to become compatible with the current case (Figure 3, A2.3). The result of this process generates a new case containing the current reaction of the interface and the confirmed problem. This new case can then be memorized in the base of case for future use (retain). The evaluator relies on his/her experience and intuition to revise the problems.

5.3 Results analysis

This phase describes the problems detected and the recommendations which results from it. Note that, for each criterion is associated a set of evaluation attributes. For example, to check the adaptation of the system to the factor luminosity, the evaluator checks the reactions of interface if the luminosity drops or increases.

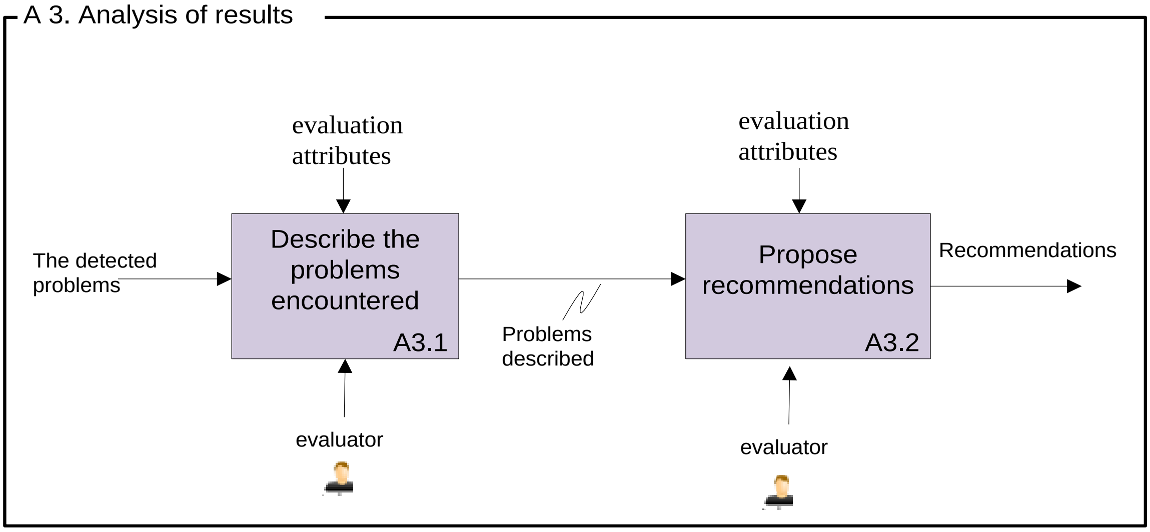

Figure 6. Analysis of the results.

The results can be communicated in the form of list of problems presented in a table. Table 1 shows the typical formalism for the description of problems, used in our method. The table includes a column “context of use” indicating the state of usage context, a column “description of the problem” containing a textual explanation of the problem and a column “consequence” indicating the possible risks of the problem.

Table 1. Formalism used to describe a problem.

Context of use	**Description of problem**	**Consequence**
State of context of use	Textual description of the problem	Textual description of consequences

The obtained results allow the realization of statistics concerning for instance the rate of success of personalized interface presentation (adequate adaptation), the average error rate, the problems most frequently encountered, etc. For example, the average error rate consists to verify if the system predictions correspond effectively to the user expectations. For this reason, we count firstly the number of erroneous predictions for each criterion. Then we calculate the error rate ER applying the following formula:

$$ER(i) = \frac{\text{nbErroneousPredictions(i)}}{\text{nbScenario}} \qquad (3)$$

With nbErroneousPrediction(i) is the number of erroneous predictions related to the criterion i and nbScenario is the number of hypothesis of use that are taken into consideration for PIS evaluation. Then, we calculate the Average Rate Error (*ARE*) of all scenarios as follows:

$$\text{ARE} = \frac{\sum \text{ER(i)}}{\text{nbCriteria}} \qquad (4)$$

With nbCriteria is the number of criteria of evaluation. Notice that if ARE is low, we may consider that the personalized system has an adapted user interface.

After problems identification, the evaluator can propose some solutions for problems detected in the previous phase. For this reason, lists containing textual descriptions of the recommendations will be prepared by the evaluator. Each list is presented in the form of a table. This table includes a column "context of use" indicating a state of the context of use (ex. use of a PDA) and a column "description of the recommendations" suggested by the evaluator (Table 2).

Table 2. Formalism used for the description of recommendations.

Context of use	**Description of the recommendations**
State of context of use	Textual description of the recommendations

These concrete descriptions of the recommendations must bring possible clear solutions, helping best the developers in the design of interfaces (Hornbaek and Frokjaer 2005). This method was applied to assess a system permitting the display of the next departures in a railway station, accessible through a web interface. The following part presents the practical application method MetSim (Method evaluation per SIMulation), dedicated for the personalized information systems evaluation.

6. Application in intelligent transport system in a logistic network

Logistics deals with the planning and control of material flows and related information in organizations, its mission is to get the right materials to the right place at the right time, while optimizing a given performance measure (e.g. minimizing total operating costs) and satisfying a given set of constraints (e.g. a budget constraint).

6.1 Presentation of the application

Transport is an integral activity of the supply chain as supports the physical flows between sites and distribution of products to customers. In fact, the logistics network can be defined as the set of successive operations of transport, to ensure the delivery of goods from point of manufacture to point of final consumption with the lowest cost and as soon as possible. The system to evaluate aims to propose the optimized itinerary to deliver goods in the distribution channel (Figure 7).. This activity is the path followed by a product from the manufacturer to the end-user. The aim of this system is to help the responsible for the activity of products distribution in their choice of itineraries according to a set of criteria (cost, distance, trip duration,..). Let's consider a set of itineraries Si (I= {1,…,n}, n is the number of possible solutions

Figure 7. Example of user interface displaying possible itineraries.

Let us remind that the principle of the evaluation by simulation is to test the application in different contexts of use. In this part, we describe an example of evaluation of the application in contexts of use in logistics domain.

6.2 Example of evaluation

The evaluation starts with the definition of the context of use. The context chosen in this evaluation is the following: the responsible for the activity of products distribution is equipped with a PDA and he/she is in a hurry. Moreover, the system is used in a closed environment (for example inside the production workshop) where the level of luminosity can be relatively low. After the context change related to the platform, the adaptation allows to change the presentation of the user interface adequately with a new characteristics of the platform (PDA having a space of display 360×240). Moreover, knowing the user agenda, the system must be able to adapt the user interface to the user in a way to present only the relevant information corresponding only to his/her activity.
In order to evaluate this adaptation, the characteristics of the containers and the components which interact will be identified by the evaluator when comparing the concrete specification of the interface before and after the adaptation. Each container or component adapted will be

described in UsiXML. Figure 8 presents the concrete specification of the interface after the process of adaptation by simulating it on a PDA.

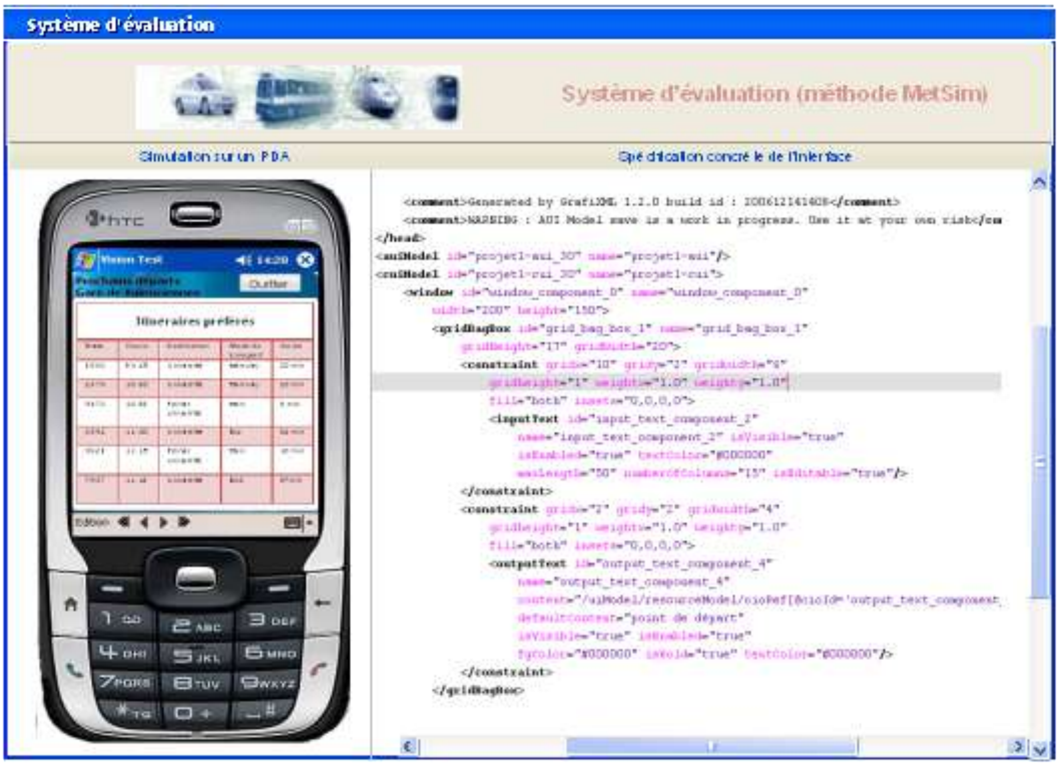

Figure 8. The concrete specification of the interface after the adaptation.

The luminosity may cause a change at the level of the colors used for the user interface, such as the passage of red characters on a black background to facilitate the vision in dark.

Figure 9. Example of the adaptations of an application to the responsable‘s speed mobility: normal mobility, normal luminosity (a) and fast mobility, low luminosity (b).

Then, the CBR uses these characteristics extracted so as to compare the reaction of the HMI of the new case with that of the former one (figure 10).

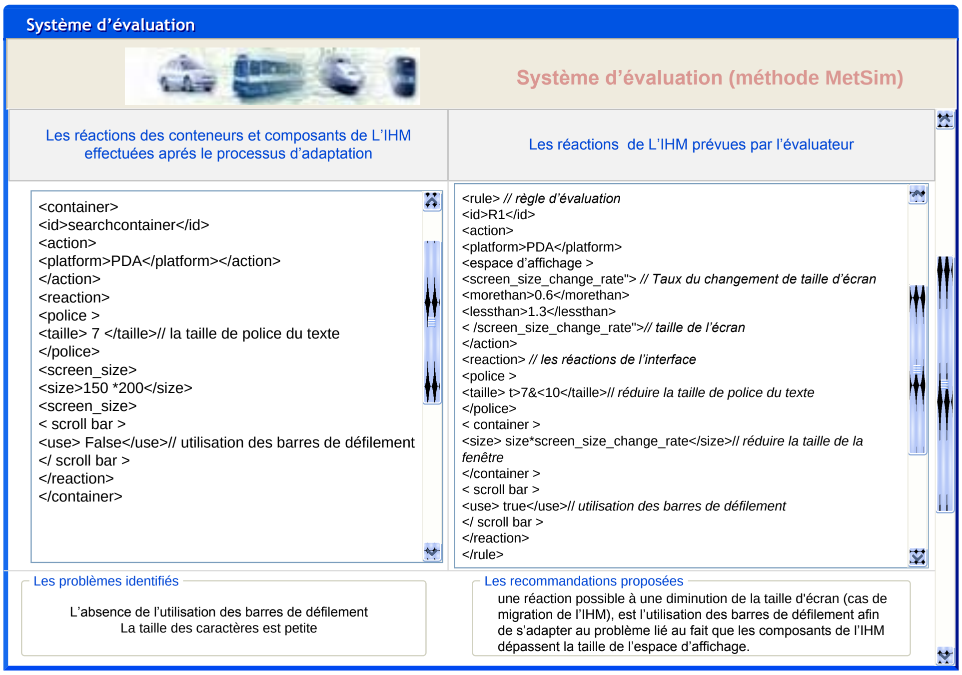

Figure 10. Real interface's reaction (left part) of the new case and reaction of the former similar one (right part).

The change of the context involved some changes at the level of the interface presentation. Among these reactions, one can quote the passage of the screen in pre- fashions lighting, display with a reduced size of screen (reduction in the size of screen).

When reminding the previous similar experiments, the evaluator extracts the most relevant case suggested problem for reuse aims; it becomes a suggested problem. It undergoes a process of revision in order to become compatible with the current case. The results of evaluation could be conveyed in the form of problems' list.

Table 3: Description of the detected problems

Context of use	**Problem's description**	**Possible consequences**
-Use of a PDA -User in a hurry -Low level of luminosity	-The size of the characters is small. -The information displayed by the interface is not clear enough.	-When using the interface displayed by the PDA, the user might encounter difficulties of interpreting the provided information, the characters being too small. -Bed legibility of information, the user can test difficulties of reading information. - A possible reaction to a fall of the level of luminosity and to the user's degree of speed when moving, is to put the characters in bold, enlarge the letters' size, space between words and lines, in addition to reducing the length of lines and the use of scrollbars.

To validate this system, an evaluation by simulation was set by evaluator. This evaluation which consists in testing the application in different contexts of use is based mainly on the average error rate. Finally, the evaluator suggested recommendations for the detected problems. Cards containing textual descriptions of the recommendations will be prepared by the evaluator (Table 4).

Table 4: Example of recommendations' description

Context of use	Description of the recommendations
Use of a PDA User in a hurry Low level of luminosity	A possible reaction to a reduction of the size of screen, is the use of the scrollbars to be adapted to the problems related to the fact that the components of the HMI exceed the size of the space of display.

7. Conclusion and perspectives

The first part of the article was centred on personalization and Personalized information system (PIS), relatively new and rich research and development domains.

We defined also the notion of interactive system evaluation; in fact an information system can be considered as an interactive system used by different people. We presented a brief review of existing methods and techniques, considered as well-known in the evaluation domain, often based on utility and usability criteria. Such methods are not directly usable for PIS evaluation. PIS evaluation is difficult and object of very few propositions and studies in the literature.

We proposed an evaluation method using simulation of a model of PIS, called MetSim (Method evaluation per Simulation). The interest of simulations is to test the application in different scenarios by varying hypothesis of use. This method does not rely on the realization of a functional prototype. It can be used to make prospective tests, to study new modes of interaction provided by a future technology. This method can be integrated within the cycle of PIS design. Indeed, this method does not disturb the user during his/her main activity (there is no necessity of explicit user intervention). It also benefits from evaluator experience in proposal of recommendation phase. Illustrations in transport domain have been given.

In order to evaluate systems more exactly and completely, several different methods should be combined. We intend to combine MetSim with other methods (questionnaire, interview…). That needs to combine data collected from MetSim and data collected from the other methods to evaluate more efficiently such PIS. Since we have tested our approach only in the logistics context, it would be also interesting to generalize the approach with other fields of application.

Acknowledgements

The present research work has been partially supported by the “Ministère de l'Education Nationale, de la Recherche et de la Technologie», the «Région Nord Pas-de-Calais», the FEDER (MIAOU, EUCUE), the ANR ADEME (Viatic.Mobilité), the PREDIM (PREDIT) (MouverPerso).

References

Ang, C.L., 1999. Enactment of IDEF0 Modes. *International Journal of Production Research*, 37(15), 3383–3397.

Askin, R.G., Ciarallo, F.W., Lundgren, N.H., 1999. An Empirical Evaluation of Holonic and Fractal Layouts. *International Journal of Production Research*, 37 (5), 961-978.

Basu, C., Hirsh, H., and Cohen, W., 1998. Recommendation as Classification: Using Social and Content-Based Information in Recommendation. In C. Rich and J. Mostow (Eds.), *Proceedings of the 15th National Conference on Artificial Intelligence*, Madison, USA, AAAI Press/MIT Press, 714-720.

Bayardo, J.R., Bohrer, W., Brice, R., Cichocki, A., Fowler, J., Helal, A., Kashyap, V., Ksiezyk, T., Martin, G., Nodine, M., Rashid, M., Rusinkiewicz, M., Shea, R., Unnikrishnan, C., Unruh, A., and Woelk, D., 1997. InfoSleuth: agent-based semantic integration of information in open and dynamic environments. In *Proceedings of the 1997 ACM SIGMOD international conference on Management of data*, 195- 206, Tucson, Arizona, United States.

Beirekdar, A., 2004. A methodology for automating web usability and accessibility evaluation by guideline. PhD Thesis, Facultés Universitaires Notre-Dame de la Paix, Namur.

Bradley, K., Rafter, R., Smyth, B., 2000. Case-Based User Profiling for Content Personalisation. *Proceedings of the International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, Lecture Notes in Computer Science, Volume 1892, pp. 62–72.

Chabrol M. Sarramia D., Techernev N., 2006, Urban traffic systems modeling methodology; *International Journal of Production Economics*, vol.99, pp. 159-176.

Chu-hua K., Christian N. M. & Chinho L., (2011), Developing global supply chain quality management systems, *International Journal of Production Research*, Vol 49, Issue 15.

Corlett, E.N., Wilson, J.R., and Corlett, N., 1995. *Evaluation of Human Work*, 2nd Edition, CRC Press.

Coyle, L., and Cunninghan, P., 2002. A Case-Based Personal Travel Assistant for Elaborating User Requirements and Assessing Offers. *Proceedings of the 6th European Conference, ECCBR 2002*, S. Craw, A. Preece (Eds.), LNAI Vol. 2416, 505-518, Springer-Verlag.

De Bra, P., Smits, D., van der Sluijs, K., Cristea, A.I., and Hendrix, M., 2010. GRAPPLE: Personalization and Adaptation in Learning Management Systems. In *ED-MEDIA 2010-World Conference on Educational Multimedia Hypermedia & Telecommunications*, Association for the Advancement of Computing in Education (AACE), June 28 - July 2, 2010 - Toronto, Canada.

Drury, C.G., 1990. Methods for direct observation of performance. In *Evaluation of Human works: a practical ergonomics methodology*, Wilson J.R. and Corlett E.N. (Eds.), Taylor & Francis.

Duchowski, A.T., 2003. *Eye tracking methodology: theory and practice*. Springer.

Fan, H., and Poole, M.S., 2006. What Is Personalization? Perspectives on the Design and Implementation of Personalization in Information Systems. Journal of Organizational Computing and Electronic Commerce, 16 (3-4), 179-202.

Feng, Y. D'Amours, S. Beauregard, R., Simulation and performance evaluation of partially and fully integrated sales and operations planning, International Journal of Production Research, 2010, VOL 48; Num 19, pages 5859-5883, ISSN 0020-7543,2010.

Francis J. (2007), Team building with the SCOR (supply chain operations reference) model, *Supply Chain Management Review*, Vol. 11, n° 2, pp. 60-65.

Ganneau, V., Demumieux, R., Calvary, G., and Demumieux, R., 2008. EMMA : Modèle Utilisateur pour la Plasticité des Interfaces Homme Machine en Mobilité. *Proceedings*

UBIMOB'08, Saint Malo, France, ACM International Conference Proceeding Series, ACM Press, 1-8.

Gowan, J. M., 2003. A multiple model Approach to personalised information access. Thesis of Master in computer science, faculty of science, University college, Dublin.

Hornbæk, K., Frøkjær, E., 2005. Comparing usability problems and redesign proposals as input to practical systems development. *Proceedings CHI 2005*, 391-400, ACM Press.

Houben, G.J., McCalla, G.I., Pianesi, F., and Zancanaro, M., 2009. *User Modeling, Adaptation, and Personalization: 17th International Conference, Umap 2009, Formerly Um and Ah, Trento, Italy, June 22-26, 2009, Proceedings*, Lecture Notes in Computer Science, Springer-Verlag Berlin and Heidelberg.

Huang S., Sheoran S., Keskar H. (2005), Computer-assisted supply chain configuration based on supply chain operations reference (SCOR) model, *Computers & Industrial Engineering*, Vol. 48, n° 2, pp. 377-394.

Huart, J., Kolski, C., Bastien, C., 2008. L'évaluation de documents multimédias, état de l'art. In Merviel S. (Ed.), *Objectiver l'humain ? Volume 1, Qualification, quantification*, Hermes, Paris, 211-250, ISBN 978-2-7462-1553-5.

ISO, 2006. *Ergonomics of human-system interaction - Part 110: Dialogue principles*. ISO 9241-110:2006.

Ivory, M., Hearst, M., 2001. The State of the Art in Automated Usability evaluation of User Interfaces. *ACM Computing Surveys*, 33 (4), 173-197.

Janssen, J., Berlanga, A.J., Heyenrath, S., Martens, H., Vogten, H., Finders, A., 2010. Assessing the Learning Path Specification: a Pragmatic Quality Approach. *Journal of Universal Computer Science*, 16 (21), 3191-3209.

Jingshan Li and Ningjian H., 2007, "Quality Evaluation in Flexible Manufacturing Systems: A Markovian Approach," *Mathematical Problems in Engineering*, vol., Article ID 57128, 24 pages, 2007.

Karat, C.M., Blom, J.O., and Karat, J.,. Designing personalized user experiences in eCommerce. Springer, 2004.

Kolodner, J., 1993. *Case-Based Reasoning*. San Mateo: Morgan Kaufmann.

Lajmi, S. Ghedira, C., Ghedira, K., 2009. CBR Method for Web-based Composition' Service. *Reading Notes in Computer Science 4879*, 314-326, Springer, ISBN 978-3-642-01349-2.

Lapena, E. L., Tobar, L. M., and Andrés, P.M.L., 2008. WebA: A Tool for the Assistance in Design and Evaluation of Websites. *Journal of Universal Computer Science*, 14 (9), 1496-1512.

Le Bodic, L., De Loor, P., Kahn, J., 2005. Umar: a modeling of multimodal artefact. In Jacko J. and Stephanidis C. (Eds.), *Proceedings of HCI International 2005*, Lawrence Erlbaum Associates.

Lieberman, H. 1995. Letizia, An Agent That Assists Web browsing. In *International Joint Conference on Artificial Intelligence* (IJCAI'95 – Montreal, Canada, August 20-25), San Francisco, USA, Morgan Kaufman Publishers.

Limbourg, Q., Vanderdonckt, J., Michotte, B., Bouillon, L., and Lopez, V. 2004. UsiXML: a Language Supporting Multi-Path Development of User Interfaces. *Proc. of 9th IFIP Working Conf. on Engineering for Human-Computer Interaction jointly with 11th Int. Workshop on Design, Specification, and Verification of Interactive Systems 199 DSV-IS'2004*, Hamburg, Lecture Notes in Computer Science, Springer-Verlag, Berlin, vol. 3425, 200–220.

MacComascaigh, M., and Andrews, W., 2009. *Leading Websites Will Use Search, Advanced Analytics to Target Content*. Gartner RAS Core Research Note.

Magoulas, G.D., Chen, S.Y., and Papanikolaou, K.A., 2003. Integrating Layered and Heuristic Evaluation for Adaptive Learning Environments. In: Weibelzahl, S. and

Paramythis, A. (Eds.), Proceedings of the Second Workshop on Empirical Evaluation of Adaptive Systems, held at the 9th International Conference on User Modeling UM, Pittsburgh, 5-14, 2003.

Mahatody, T., Sagar, M., Kolski, C., 2010. State of the Art on the Cognitive Walkthrough method, its variants and evolutions. *International Journal of Human-Computer Interaction*, 26 (8), 741-785.

Nielsen, J., 1993. *Usability engineering*. San Diego, CA: Academic Press.

Nielsen, J., and Molich, R., 1990. Heuristic evaluation of user interfaces. *CHI'90 Conference Proceedings*, Seatle, ACM Press, New York, 349-356.

Nodine, M., Ngu, A.H.H, Cassandra, A. and Bohrer, W.G., 2003. Scalable Semantic Brokering over Dynamic Heterogeneous Data Sources in InfoSleuth™. *IEEE Transactions on Knowledge and Data Engineering*, 15(5), 1082-1098.

Norsham, I., Norazah, Y., and Puteh, J., 2009. Adaptive Course Sequencing for Personalization of Learning Path Using Neural Network. *Int. J. Advance. Soft Comput. Appl.*, 1 (1),

Payne J.and Cariapa V., "A fixture repeatability and reproducibility measure to predict the quality of machined parts," *International Journal of Production Research*, vol. 38, no. 18, pp. 4763–4781, 2000

Pazzani, M., Muramatsu, J., and Billsus, D., 1996. Syskill & Webert: Identifying interesting web sites. *Proceedings of the National Conference on Artificial Intelligence*, Portland, OR.

Piller, F., and Tseng, M., Handbook of Research in Mass Customization and Personalization. World Scientific Publishing Company, 2010.

Sancho, P., Martínez, I., and Fernández-Manjón, B, 2005. Semantic Web Technologies Applied to e-learning Personalization in <e-aula>, *Journal of Universal Computer Science*, 11 (9), 1470-1481.

Sears, A., and Jacko, J.A., 2008. *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies (2nd Ed.)*. CRC Press, New York, United States.

Senach, B., 1990. *Evaluation ergonomique des interfaces Homme-Machine : une revue de la littérature*. Research report, INRIA, n°1180, Sophia Antipolis, France.

Soui, M., Abed, M., Kolski, C., and Ghédira, K., 2007. Criteria devoted to evaluate personalized interactive systems. In M. Abid, A. Hadj Kacem, M. Jmaiel and M. Lahiani (Eds.), *Nouvelles tendances technologiques en génie électrique & informatique*, GEI'2007, Edition CPU, Sfax, 119-125, ISBN 978-9973-61-631-9.

Soui, M., Abed, M., Kolski, C., and Ghédira, K., 2010. Evaluation by simulation for personalized information systems. *8th International Conference of Modeling and Simulation, MOSIM'10 "Evaluation and optimization of innovative production systems of goods and services"* (May 10-12, 2010), Hammamet, Tunisia.

Sugimori Y., Kusunoki, Cho F., Uchikawa S., 2007, Toyota production system and Kanban system Materialization of just-in-time and respect-for-human system, , *International Journal of Production Research*, Vol. 15, No.6, 553-564.

Tarby, J-C., Ezzedine, H., and Kolski, C. 2009. Prevision of evaluation by traces during the software design of interactive systems: two approaches compared. In A. Seffah, J. Vanderdonckt, and M. Desmarais (Eds.), *Human-Centered Software Engineering: Architectures and Models-Driven Integration*, Springer HCI Series, 257-276.

Vanderdonckt, J., 1999. Development Milestones towards a Tool for Working with Guidelines. *Interacting with Computers*, 12 (2), 81–118.

Whitefield, A., Wilson, F., and Dowell, J., 1991. A framework for human factors evaluation. *Behaviour & Information Technology*, 10 (1), 65-79.

Yang C. L., Huang C. L. and Chuang S. P. 2011 : Outsourcing evaluation system based on AHP/ANP approach for LED industry, Journal of Statistics & Management Systems, Vol 14 Number, ISSN 0972-0510.

Zimmermann, A., and Lorenz A., 2008. LISTEN: a user-adaptive audio-augmented museum guide. *User Modeling and User-Adapted Interaction*, 18 (5), 389-416.